DOCUMENT RESUME

ED 249 507 CS 208 600

AUTHOR	Escoe, Adrienne
TITLE	Writing a Writing Assessment: Saying What You Want to Say Isn't as Simple as It Seems.
INSTITUTION	Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	SWRL-TR-79
PUB DATE	12 Dec 82
CONTRACT	NE-0-00-3-0064
NOTE	54p.
PUB TYPE	Reports - Descriptive (141) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Educational Assessment; *Evaluation Criteria; Evaluation Methods; *Job Skills; Secondary Education; *Test Construction; *Test Format; *Writing (Composition); *Writing Evaluation; Writing Skills

ABSTRACT

Since acceptable writing is essential to success in job training programs and in many entry-level jobs, a writing sample was included in the Training and Employment Prerequisites Survey, a multiple-choice test about skills like mechanics, usage, and spelling. The two writing prompts asked students to give directions for finding a location in a school building (Form A) and to compose a business letter that requests adjustment of an order (Form B). The scoring criteria for Form A were as follows: (1) submits scorable sample, (2) includes important information about direction of movement, (3) includes critical information about features along the way, (4) writes directions in logical order, (5) uses precise language, (6) uses correct grammar and complete sentences, (7) capitalizes and punctuates correctly, and (8) spells correctly. Criteria for Form B took into consideration how well the student stated the problem and the desired course of action, used precise language, used language appropriate for a business letter, excluded unrelated sentences, used correct grammar and complete sentences, capitalized and punctuated correctly, and spelled correctly. Field tests of the two forms yielded some changes in criteria and the language of the prompts. Both writing sample assessments will be subjected to further revision based on additional experience with them. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

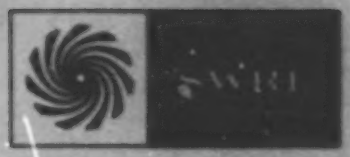
ED249507

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.



SWRL EDUCATIONAL RESEARCH AND DEVELOPMENT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

SWRL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Writing a Writing Assessment: Saying What You Want to Say Isn't as Simple as It Seems

CS 208 000

BEST COPY AVAILABLE

This document has been distributed to a limited audience for a limited purpose. It is not published. Copies may be made only with the written permission of SWRL Educational Research and Development, 4665 Lampson Avenue, Los Alamitos, California 90720. The work upon which this document is based was performed pursuant to Contract NE-0-00-3-0064 with the National Institute of Education. SWRL reports do not necessarily reflect the opinions or policies of the sponsors of SWRL R&D.

SWRL EDUCATIONAL RESEARCH AND DEVELOPMENT

TECHNICAL REPORT 79

December 12, 1982

WRITING A WRITING ASSESSMENT: SAYING WHAT YOU WANT TO SAY ISN'T AS SIMPLE AS IT SEEMS

Adrienne Escoe

ABSTRACT

The paper presents an ethnographic account that traces the creation of a writing sample assessment and provides guidance for those who design and develop assessments of students' writing.

WRITING A WRITING ASSESSMENT: SAYING WHAT YOU WANT TO SAY ISN'T AS SIMPLE AS IT SEEMS

Adrienne Escoe*

School districts have begun, in recent years, to devote more resources (e.g., staff, materials) to writing instruction. Many have developed programs of instruction centered on discourse products. Typically, students are given a sample prompt, they write a response, and then have their writing evaluated. We've seen collections of prompts used as the foundation of instruction; that is, a new prompt is assigned each week, every two weeks, or some other time interval (e.g., Fullerton Union High School District Competency Review: Paragraph Review Booklet, 1981). Usually teachers evaluate ("grade") students' papers, sometimes place comments on them, return the papers, and discuss a handful or so.

Evaluation criteria almost always include grammar and usage, spelling, capitalization, punctuation--all, elements of form. Everybody--teachers, parents, employers--wants good writing form. And most evaluations of form are easy to come by. It's not too difficult to set standards (i.e., criteria) for correct form: either the first word of a sentence is capitalized or it's not; either the period's at the end or it's not. Establishing standards for <u>content</u>, however, is a different story.

---

*Students and teachers from the Los Angeles Unified School District and Garden Grove Unified School District participated in the study, and the following SWRL staff were involved in various aspects including creating the original versions of the writing prompts, trying them out informally, serving as field-test readers and liaison with the teachers, and processing data: Barry Bachelor, Bruce Cronnell, Larry Gentry, Ann Humes, Joe Lawlor, Shirley Paddit, Jan Perkins, Gail Reneau, Roger Scott, Karen Smith, David Snow, and Nancy Yaman.

Characteristics of content, e.g., staying to the main topic, are included in some criteria used for assessing written discourse; more often, the standards are broad and rather vague, with the only criterion often being whether the writer has responded directly to the prompt given.

Writing a good writing prompt is not so easy; but harder still is seeing to it that the criteria and the prompt are written to suit each other well, establishing criteria of form and content, and then checking empirically to make sure that students respond to the prompt and that teachers evaluate the writing samples as intended. That's the only way thus far we know how to determine whether the criteria and the prompt are good, i.e., the prompt is productive and the sample is judgeable. And without those characteristics a writing sample is not worth the paper it's written on.

This paper is an ethnographic account of the creation of a writing sample assessment and is a guide for those who design and develop assessments of students' writing.

To include a Writing Sample

The occasion for the assessment was the development of instrumentation to ensure that individuals who are popularly termed "remedial" and "hard-to-employ" acquire the prerequisites for success in training and employment. Through surveys and analyses of instruction and of job-training prospects for these individuals, we determined that clear, correct writing was a desired and teachable accomplishment. And one way to estimate students' writing accomplishments is to have them choose correct responses to

multiple-choice items about skills like mechanics, usage, and spelling. We included items of this sort on the TEPS (the acronym for Training and Employment Prerequisites Survey). But multiple-choice items do not describe productive writing, i.e., written discourse. Our thinking was that high-risk students (those headed for unemployment, unstable employment, or marginal employment) needed to be able to produce acceptable writing to enhance their chances of sidestepping a predictable fate; acceptable writing is essential to success in job-training programs and in many entry-level jobs. Thereupon, we decided to include writing samples in the TEPS.

Writing the TEPS Writing Sample Assessments

The next stage was to ask questions: What kind of writing would entry-level employees or workers-in-training need to do? What kind of writing had they practiced in school? What kind of writing would "get at" the skills we wanted to describe? For answers we looked to SWRL research on writing. Several documents were particularly helpful (Escoe, 1982a,b; Gentry, 1982; Humes, 1980; Humes, Cronnell, Lawlor, Gentry, & Fieker, 1980; Nack, 1982).

Training and employment data indicate that clear, accurate communication is one of employers' top priorities. And service to the public is involved in the majority of entry-level positions. So it was a short step to specify a writing sample whose primary emphasis was to communicate a message clearly and serve the needs of other people. But other considerations were equally important, among them, topic familiarity, writing type practiced in school, and simple vocabulary.

To demonstrate fairly students' accomplishments, it was essential that the writing sample assessment consist only of tasks that would be familiar, that is, practiced in school. One skill area emphasized in remedial reading and English classes, the data indicate, is sequential information. Students often read sequentially-arranged text and answer questions about the order of events. Other frequent activities in these classes are writing statements in logical sequence and developing a paragraph the same way. Another common focus of instruction is interpreting and writing business correspondence. Typically, programs stress written conventions, such as format and appropriate language, though of the two examples, letter format is more commonly taught. But since multiple-choice items that assessed knowledge of letter format were already designed into the TEPS, and business-like writing is so important a part of employment success, appropriate language in a business letter was chosen for inclusion in the TEPS writing sample assessment.

Next, the vocabulary and syntax of the assessment prompt and the directions for students had to be familiar and simple. For vocabulary, I referred to the EDL core list, grades three to seven for Form A, grades six to nine for B. In that way, we had some assurance that students had encountered words in instructional materials or at least in widely used test materials such as the California Test of Basic Skills. Then, to effect syntactic simplicity, sentences were written to be straightforward, i.e., without troublesome transformations and inclusions. The idea was to avoid any characteristic of the assessment likely to present reading difficulties that would interfere with writing performance.

To assess a student's writing performance fairly, the prompt and the directions must clarify what is expected. For example, if correct spelling is one of the skills to be assessed, students should be made aware of that fact; otherwise, some students may not attend to spelling, under the erroneous assumption that spelling "doesn't count."

At this point, three additional general criteria were established for developing a writing sample assessment: produces 1) easily-scored writing, 2) diagnostic information, and 3) reliable information. But measuring the assessment against these three had to wait until the previously described criteria were worked out. To sum up, the general criteria and, correspondingly, the specific criteria for the TEPS were the ones shown on the following page.

The next step in writing the writing sample prompt involved creativity more than anything else. Thumbing through reading and English workbooks and published and district assessments, and SWRL technical documents, but mostly just brainstorming, we listed, evaluated, and ultimately rejected scores of topics for the writing samples. Two topics, however, tentatively seemed to conform to all the established criteria: giving directions for finding a location in a school building (Form A) and composing the body of a business letter that requests adjustment of an order (Form B). Both topics seemed likely to elicit a discourse product whose primary emphasis would be to communicate a message clearly and serve a person's needs. Directions for getting to a room in a school building had to be clear, accurate, and given in a logical sequence to be helpful to the hypothetical person who followed them. The business letter had to be written clearly with desired action stated precisely so that the imaginary recipient

Writing Sample Prompt

General Criteria	TEPS Criteria	
	Form A	Form B
1. Skills valued for training and entry-level employment	Clear, accurate communication Public service activity	
2. Skill areas practiced in school	Sequential information	Business correspondence
3. Format practiced in school	Writing in sequence of occurrence	Writing a business letter
4. Simple, familiar vocabulary	Core vocabulary grade levels 3 through 7	Core vocabulary grade levels 6 through 9
5. Simple, straightforward syntax	Complex sentence constructions avoided	
6. Precise statement of requirements	All criteria for evaluation spelled out to examinees	
7. Produces easily-scored writing	Generates scoring criteria that require little or no training of evaluators, and that pertain specifically to the prompt	
8. Produces diagnostic information	Generates scoring criteria that separately pertain to features of content and form	
9. Produces reliable information	Readers tend to agree in their assessments	

would be able to respond satisfactorily to the request, i.e., make a satisfactory adjustment to the order. Each topic, further, represented practiced skills and skill formats: writing events in sequence, and writing a business letter. Vocabulary of the writing prompt could be kept familiar and simple. For the Form A prompt, no technical terms would be needed; and students should be quite familiar with the names of features of a school building and simple directional terms (e.g., left, right, north, south). For Form B, some familiarity with the business world would be needed, so we looked long and hard for a type of business that most students would know about, and one for which knowledge of technical terms was not required. After countless attempts, we zeroed in on a business that handled office supplies. Even if students were completely ignorant of the office supply business, they should be able to respond to this topic because of school and home uses of office supplies and because of the open nature of the prompt itself (which will be discussed later). Moreover, since this topic was a candidate for the Form B assessment, our thinking was that the more advanced students (Form B examinees) would tend to be closer to entering the job market. And, realistically, if they had no knowledge <u>at all</u> of common materials used in an office, they probably wouldn't be ready for entry-level jobs or job training (even the <u>slightest</u> knowledge of materials would get students by with the prompt as it came to be written). The next consideration in writing the prompt was syntax. Nothing more than careful construction was required here--no complicated transformations, no intrusions, no very long sentences. Next, students had to be told precisely what teachers would be "looking for" in their writing. This

criterion was difficult to meet; only through a series of field tests that included reworded prompts were we able to advise students on what to include in their writing without having them produce virtually literal recasting of the prompts. Our experience in developing the TEPS writing sample prompts confirmed the observation by Humes et al. that piloting with target populations is essential (1980).

Likewise, the seventh general criterion can be addressed only by field testing: readers (ideally those who represent readers who will function under actual conditions of TEPS administration and scoring, i.e., teachers) must actually use scoring criteria to evaluate students' writing. These readers must also be encouraged to comment upon the ease (or difficulty) of scoring the samples.

Referring to the eighth criterion, diagnostic information can be provided only when the scoring criteria require attention to the specific content features of the writing task, rather than to the general writing performance or features of form alone. With only general performance assessed, knowledge gained of students' writing is unproductive for planning instruction; with only form features assessed, the form of performance may influence the assessment of content features, thus contaminating potential diagnostic information.

Finally, the writing sample prompt combined with scoring criteria should yield evaluations that are stable across readers of the writing sample. Understand that no one is suggesting perfect agreement among raters; rather, evaluations should be reliable enough so they are unlikely to be idiosyncratic to a single reader (i.e., teacher). Teachers' judgments based on long-term contact with their own students' writing are

bound to be more useful for classroom instruction than any one-shot assessment; but an advantage of using a survey across teachers lies in inter-classroom activities, such as informing next-term teachers of students' writing accomplishments. With each refinement of the TEPS writing sample assessments, we gathered data on interrater reliability: we used classroom teachers as well as SWRL professionals to evaluate students' writing samples. Resulting data were analyzed and thus recommended further refinements. The remainder of this paper is a description and narrative of each major revision of the two TEPS writing sample assessments as they evolved through a series of field tests.

Evolution of the Writing Sample Assessment, Form A

The assessment included the following note to the student, prompt, and scoring criteria:

TRAINING AND EMPLOYMENT PREREQUISITES SURVEY

ENGLISH LITERACY I

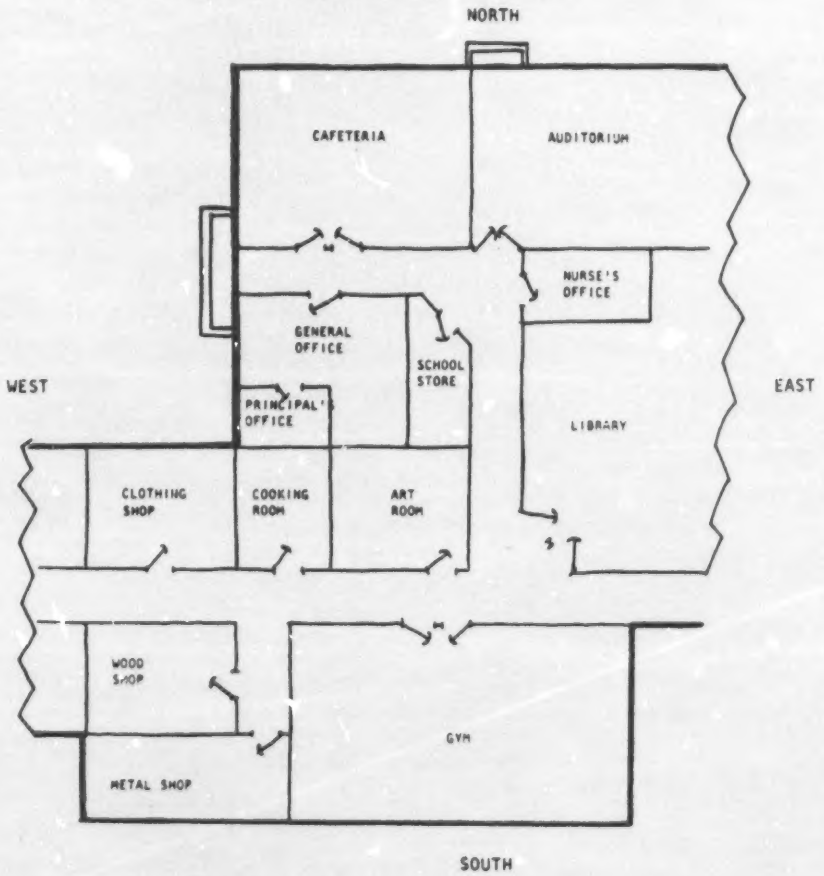
WRITING SAMPLE

NOTE TO THE STUDENT

> The purpose of this writing sample is to find out what you have learned about writing. Write your name at the top of the lined paper. Then read the directions on the other side of this page. Keep the directions in front of you to remind you of the things to include when you write. After you have finished, your teacher will collect your paper and this page. You may begin.

DIRECTIONS: Study the map of part of a school building. Locate the metal shop and the cafeteria. Write directions that tell a new student how to get from the metal shop to the cafeteria.

- Write the directions in an order that is easy to follow.
- Include all important information about which way to go.
- Include all important information about what to watch for along the way.
- Be sure to use exact words.



WRITING SAMPLE Scoring Criteria

	Acceptable	Unacceptable
CONTENT:		
1. Submits scorable sample.	___	___
2. Includes important information about direction of movement.	___	___
3. Includes critical information about features along the way.	___	___
4. Writes directions in logical order.	___	___
5. Uses precise language.	___	___
6. Limits the paragraph to one main idea.	___	___
FORM:		
7. Uses correct grammar and complete sentences.	___	___
8. Capitalizes and punctuates correctly.	___	___
9. Spells correctly.	___	___
10. Writes legibly with appropriate margins and indentation.	___	___

Before administering these materials to the first group of students, we changed the scoring criteria to more accurately address the features of the particular writing task; that is, limiting the paragraph to one main idea (criterion six) was inappropriate for a writing sample that provides sequential directions. We also removed criterion one (Submits scorable sample.) from the CONTENT category, because scorability includes features of both content and form. Then, criterion ten (Writes legibly with appropriate margins and indentation.) was eliminated. Its function

as a criterion for assessing legible writing was already claimed by criterion one (Submits scorable sample.); furthermore, assessing margin formation was considered too subjective for easy or reliable scoring. The remaining utility of the criterion (assessing indentation) was simply not great enough for the time spent evaluating this feature in students' writing.

Upon the recommendation of a SWRL model for evaluating writing samples (Humes, 1980), a third rating was added to the scoring system: "Good." So as we headed for the first field test, criteria for scoring the samples looked like this:

WRITING SAMPLE Scoring Criteria

	Good	Acceptable	Unacceptable
1. Submits scorable sample.	___	___	___
CONTENT:			
2. Includes important information about direction of movement.	___	___	___
3. Includes critical information about features along the way.	___	___	___
4. Writes directions in logical order.	___	___	___
5. Uses precise language.	___	___	___
FORM:			
6. Uses correct grammar and complete sentences.	___	___	___
7. Capitalizes and punctuates correctly.	___	___	___
8. Spells correctly.	___	___	___

The last step prior to administering the assessment was to expand the description of each criterion to include characteristics of each rating (Good, Acceptable, Unacceptable) for each criterion. The following directions* accompl'shed this task:

DIRECTIONS FOR SCORING THE WRITING SAMPLE - A

Teachers are asked to score the writing sample which students completed on the lined paper. Guidelines for scoring the writing sample are presented below. You should do your scoring directly on the back of students' answer sheets in the spaces numbered 121-125. Use a number 2 pencil to do your scoring. The scoring guide below describes three performance ratings for each criterion: good, acceptable, and unacceptable. Fill in the space marked A for good, B for acceptable, and C for unacceptable.

Most teachers will only need to read the scoring guidelines a few times before they remember the criteria for each performance rating. A simple scoring aid is provided after the guidelines.

Scoring Guide (121 through 128)

121. Submits scorable sample.

Good: (This rating is not used for this criterion.)

Acceptable: Has both of the following:

a. legible writing
b. writing relevant to directions in prompt

Unacceptable: Has any one of the following:

a. illegible writing
b. writing not relevant to directions in prompt
c. no response

CONTENT:

122. Includes important information about direction of movement.

Good: Includes at least four of the following directional segments:

a. (from metal shop): out or straight ahead or down the hall or north
b. right or east
c. left or north
d. left or west
e. right or north

Acceptable: Includes all of the three major directional segments as follows:

a. (at first "T" in hallway): right or east
b. left or north
c. left or west

Unacceptable: Does not include all of the three major directional segments as follows:

a. (at first "T" in hallway): right or east
b. left or north
c. left or west

*Directions refer to criteria numbered 121-128. These three-digit numbers correspond to the Answer Sheet, where ratings were recorded. In fact, 121 refers to criterion one, 122 to criterion two, and so forth. Note also that the designation of the Form was changed from "1" to "A" to conform to the designations of the TEPS Pre-Algebra (Forms A and B).

BEST COPY AVAILABLE

BEST COPY AVAILABLE

123. Includes critical information about features along the way.

Good: Includes at least five of the following features:

a. metal shop
b. hall or hallway*
c. cooking room
d. art room
e. gym
f. library
g. school store
h. nurse's office
i. auditorium
j. general office
k. cafeteria

*NOTE: "hall" or "hallway" can be counted more than once if more than one hallway is mentioned (for example, "the first hallway" and "across the hallway from the general office")

Acceptable: Includes three or four features (see list of features under Good, above).

Unacceptable: Includes no or one or two features (see list of features under Good, above).

124. Writes directions in logical order.

Good: All information is given in order of movement.

Acceptable: Information is given mostly in order of movement (one or two directional segments and features are mentioned out of order of movement).

Unacceptable: Information is given mostly in scrambled order. The order written would make it difficult for a new student to follow the directions.

125. Uses precise language.

Good: Precise words are used in most or all cases to describe directions, for example:

"cooking room" vs. "room"
"hallway" vs. "place"
"left" or "north" vs. "that way"
"cafeteria" vs. "there"

Acceptable: General words are used in most cases to describe directions; however, precise words may be used in a few cases. (See examples under Good, above.)

Unacceptable: Words used are so general that a new student would find it difficult to follow the directions.

FORM:

126. Uses correct grammar and complete sentences.

Good: Has both of the following:

a. few or no grammatical errors
b. no incomplete sentences (fragments, run-ons)

Acceptable: Has any one of the following combinations:

a. few or no grammatical errors and few incomplete sentences
b. some grammatical errors and no incomplete sentences
c. some grammatical errors and few incomplete sentences

Unacceptable: Has both of the following:

a. many grammatical errors
b. many incomplete sentences

127. Capitalizes and punctuates correctly.

Good: Has few or no capitalization and punctuation errors.

Acceptable: Has some capitalization and punctuation errors.

Unacceptable: Has many capitalization and punctuation errors

128. Spells correctly.

Good: Has few or no misspelled words.

Acceptable: Has several different* misspelled words.

Unacceptable: Has many different* misspelled words.

Scoring Aid (121 through 128)

Place the student's answer sheet under this page of the Manual, aligning rows 121 through 128 on the answer sheet with 121 through 128 below:

(A) Good
(B) Acceptable
(C) Unacceptable

121. Submits scorable sample.

122. Includes important information about direction of movement.

123. Includes critical information about features along the way.

124. Writes directions in logical order.

125. Uses precise language.

126. Uses correct grammar and complete sentences.

127. Capitalizes and punctuates correctly.

128. Spells correctly.

*More than one instance of misspelling the same word should be evaluated as one misspelled word.

BEST COPY AVAILABLE

Participants in the field test for TEPS, English Literacy, Form A Writing Sample were 18 seventh-grade students (13 boys, 5 girls) who attended remedial English classes in an urban junior high school, their four English teachers,* and two SWRL staff. Students were given the prompt and the note shown earlier. They wrote their responses on lined white paper, some using pencil, but most, pen. All students had Spanish surnames, but their proficiency in spoken English could not be determined from these data. The results obtained from the field test included students' performances, reliability of readers' (teachers and SWRL staff) ratings and their commentary (readers' comments for this and subsequent field tests are included in a SWRL report, Development of the TEPS Writing Sample Assessment). Students' performance by criterion and estimates of interrater reliability follow.

Summary of Students' Performance By Criterion and Rater**

Criterion	Rating											
	Good(A)				Acceptable(B)				Unacceptable(C)			
	R1**	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
1.	2	4	0	0	14	14	18	15	0	0	0	3
2.	6	7	9	10	5	7	3	1	5	4	6	7
3.	9	7	10	8	4	7	7	4	3	4	1	6
4.	12	16	8	11	1	0	8	4	3	2	2	3
5.	7	9	18	15	7	8	0	1	2	1	0	2
6.	0	0	0	0	7	10	0	5	9	8	18	13
7.	2	1	0	1	7	8	0	5	7	9	18	12
8.	4	5	7	5	8	10	7	7	4	3	4	2

*Four teachers divided students' papers, so each pair of teachers read half the papers. For computations, the four teachers were treated as two raters.

**Raters 1 and 2 are junior high school (seventh-grade) English teachers. Raters 3 and 4 are SWRL staff.

***Rater 1 omitted ratings for two students; hence the total number of ratings for Rater 1 is 16.

An examination of the data above shows a natural clustering of teachers' ratings and SWRL staff ratings. All in all, the teachers tended to rate more performances good or acceptable than the SWRL staff members; however, upon closer study, the data show just the opposite for criterion five (Uses precise language.). This peculiarity prompted a rating sequel that involved two other SWRL readers and the same student responses. The results of this small-scale rating activity are described later. Except for criterion two (Includes important information about direction of movement.) and possibly criterion three (Includes critical information about features along the way.), these coefficients were discouraging.

Summary of Estimated Interrater Reliability by Criterion
Between Pairs of Four Raters
(Pearson Correlation Coefficients)

	Raters*					
Criterion	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
1.	.1336	**	.0000	**	-.1195	**
2.	.6741	.6362	.7724	.7699	.6458	.8081
3.	.5143	.6192	.4883	.5447	.6374	.6340
4.	.2936	.5430	.0186	.1768	-.2579	.2553
5.	.3343	**	.0513	**	.6029	**
6.	.4535	**	-.0814	**	.3051	**
7.	.2446	**	.0377	**	.4891	**
8.	.3280	.3440	.0397	.8484	.7755	.7330

*Raters 1 and 2 are junior high school (seventh-grade) English teachers. Raters 3 and 4 are SWRL staff.

**Coefficient could not be computed because variance was zero.

So we looked at results for individual students. And there it hit us. If "Good" ratings were collapsed with "Acceptable" ones, agreement between pairs of raters looked as if it would skyrocket. This observation was set aside temporarily while the Form B assessment was field tested; it too yielded such results. During this time, readers' comments about the Form A assessment were reviewed.

Based on student performance, estimated interrater reliability, and readers' comments, the note to the student, the prompt (including the diagram), and the DIRECTIONS FOR SCORING THE WRITING SAMPLE - A* (now included within a draft of the administration manual, under sections "Administration: Writing Sample" and "Directions for Describing the Writing Sample - Form A Guide and Form A Summary of Descriptions") were revised as follows (the Note to the Student was now designated as Form A rather than Form 1):

TRAINING AND EMPLOYMENT PREREQUISITES SURVEY

ENGLISH LITERACY A

WRITING SAMPLE

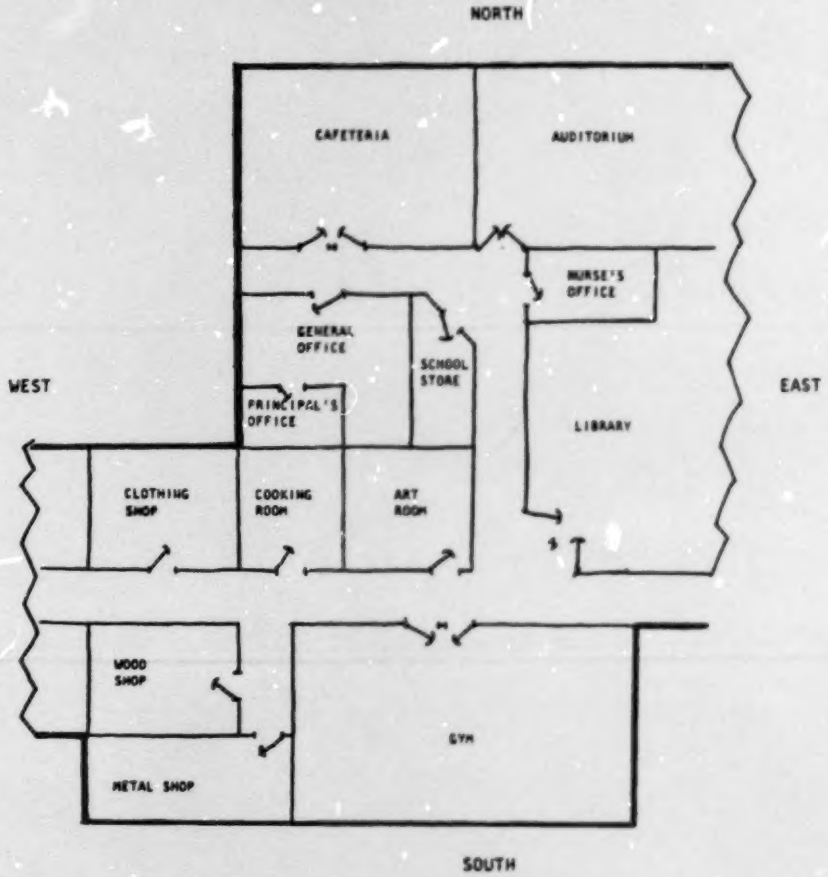
NOTE TO THE STUDENT

The purpose of this writing sample is to find out what you have learned about writing. Write your name at the top of the lined paper. Then read the directions on the other side of this page. Keep the directions in front of you to remind you of the things to include when you write. After you have finished, your teacher will collect your paper and this page. You may begin.

*Directions refer to criteria numbered 111-118. These three-digit numbers correspond to the Answer Sheet, where ratings were recorded. In fact, 111 refers to criterion one, 112 to criterion two, and so forth.

DIRECTIONS: Study the map of part of a school building. Locate the metal shop and the cafeteria. Write a paragraph that tells a new student how to get from the metal shop to the cafeteria.

- Write the directions in an order that is easy to follow.
- Include all important information about which way to go.
- Include all important information about what to watch for along the way.
- Be sure to use
  - exact words
  - good grammar and complete sentences
  - correct punctuation and spelling



Administration: Writing Sample

The Writing Sample is the part of the English Literacy Survey that offers an opportunity to describe characteristics of an individual's written response to a prompt. Characteristics described include content (ideas represented) and form (language usage and mechanics).

- The Writing Sample is not intended to be a timed activity. Within reasonable limits, each examinee should be given enough time (about 15-25 minutes) to complete the item.
- Answer Sheets and Survey Booklets should not be available to examinees during the administration of the Writing Sample.
- Distribute to each examinee the single sheet marked WRITING SAMPLE (Form A or B) and a sheet of lined paper.
- Have examinees read the information in the box below the words NOTE TO THE STUDENT. Examinees should be able to read independently and follow the directions for the writing sample. Be sure each examinee has written her or his name on the lined paper.
- After examinees finish writing, collect the single sheet of directions for the WRITING SAMPLE and the actual writing sample.

Directions for Describing the Writing Sample (Forms A and B)

Teachers are asked to describe the writing sample which students completed on the lined paper. Guidelines for describing the writing sample are presented below. You should use students' answer sheets and fill in the spaces enclosed in the box near the right bottom corner of SIDE ONE. This box will be marked TEACHER SCORED WRITING SAMPLE.

The guides below (Forms A and B) describe student performance for several characteristics. Refer to the guides to fill in the bubble whose description comes closest to the student's performance for each characteristic. A one-page Summary of Descriptions follows the Guide for each form. The Summary can simplify the task of describing characteristics.

Form A Guide. Spaces numbered 111 through 118 are used to describe characteristics of writing samples for Form A. Use the descriptions below to fill in the bubbles on the student's answer sheet. Fill in only one bubble for each number. Do not fill in any bubbles if the sample is any of the following:

- illegible
- irrelevant to directions given
- not written

		(A)	(B)	(C)	(D)
111.	Estimates size of writing sample (based on 8 1/2 x 11 sheet of paper).*	Writes about a full page or more.	Writes about 3/4 page.	Writes about 1/2 page.	Writes about 1/4 page or less.

*NOTE: If handwriting is much larger or smaller than average, judge sample size according to what it would be if handwriting were of average size.

CONTENT:

112. Includes important information about direction of movement.

Includes all of the three major directional segments:

- right or east (at first "T" in hallway)
- left or north
- left or west.

(A)	(B)
Yes	No

113. Includes critical information about features along the way.

Includes three or more of the following features:

metal shop	hall or hallway*
cooking room	art room
gym	library
school store	nurse's office
auditorium	general office
cafeteria	door or doorway

*NOTE: "hall" or "hallway" can be counted more than once if more than one hallway is mentioned (for example, "the first hallway" and "across the hallway from the general office").

(A)	(B)
Yes	No

114. Writes directions in logical order.

All or most information is given in order of movement.

(A)	(B)
Yes	No

115. Uses precise language.

Uses precise words to describe directions, for example:

"cooking room" or "room" (rather than "place"

"left" or "north" (rather than "that way")

(A)	(B)
Yes	No

FORM:

116. Uses correct grammar and complete sentences.

The total number of grammatical errors and incomplete sentence* errors is _____.

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: Count one error each time a segment of a run-on sentence should have been a separate sentence.

117. Capitalizes and punctuates correctly.

The total number of capitalization and punctuation errors is _____.*

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: Do not count as errors the places where a run-on sentence should have been segmented (and thus capitalized).

118. Spells correctly.

The total number of spelling errors is _____.*

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: More than one instance of misspelling the same word the same way should be evaluated as just one misspelled word.

Form A Summary of Descriptions. Place the student's answer sheet next to this page of the directions, to help you fill in the answer sheet quickly.

111.	Sample size	(A) full page +	(B) 3/4	(C) 1/2	(D) 1/4 or less	
112.	Direction of movement	(A) (three segments)	(B)			
		Yes	No			
113.	Critical features	(A) (three features +)	(B)			
		Yes	No			
114.	Logical order	(A)	(B)			
		Yes	No			
115.	Precise language	(A)	(B)			
		Yes	No			
116.	Errors in grammar and incomplete sentences	(A)	(B)	(C)	(D)	(E)
		0-1	2	3	4	5+
117.	Errors in capitalization and punctuation	(A)	(B)	(C)	(D)	(E)
		0-1	2	3	4	5+
118.	Errors in Spelling	(A)	(B)	(C)	(D)	(E)
		0-1	2	3	4	5+

Notice that the prompt now spelled out the requirement for exact words, good grammar and complete sentences, and correct punctuation and spelling. Notice too that symbols for two exterior (and extraneous) doors were removed. A minor adjustment was made in a line that represents a wall of the principal's office--previously that line ran through the letter s in principal's. Another minor change was that the form of the TEPS mentioned in the heading of the note to the student was changed from 1

to A to correspond to the Pre-Algebra component. The revisions made upon the readers' directions, however, were major.

It occurred to us that readers may have been scoring the writing samples according to habit, rather than attending to descriptions associated with individual criteria; choosing among ratings called good, acceptable, and unacceptable, and recording these as A, B, and C seemed as if they could reinforce customary paper-grading practices. Since the intent of the TEPS was to describe student accomplishments and not so much to rate or grade them, a solution we decided to try was a descriptive rating system. In addition, individual criteria were revised. The highlights of these revisions were as follows: readers were directed to set aside papers that were unscorable (according to characteristics given) rather than darken a bubble for a criterion; the size of the writing sample was to be estimated (to eliminate the vague, judgmental descriptions for the form criteria, e.g., few, some, many, and several, by directing readers to count errors for these criteria--school systems could set their own error standards based on size of sample); and content criteria could be rated only as present in the sample or not present (yes or no). Also, each form criterion was followed by an annotation that gave raters additional guidance (especially in dealing with run-on sentences).

Two SWRL staff used the revised directions to describe the same writing samples. Without a group of student papers that were in response to the new prompt, of course, new data would be limited to new rating directions and new raters. A similar wording change for the Form B prompt, however, was field tested and did improve the results.

The following summary of students' performance and interrater reliability estimates reflects revised criteria and ratings from two SWRL staff who hadn't before seen or heard about (I was assured) any of the materials involved in the writing sample project:

Summary of Students' Performance
By Criterion and Rater (Revision)

	Rating*									
	A		B		C		D		E	
Criterion	R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
1.	0	0	0	0	7	7	11	11		
2.	11	13	7	5						
3.	16	15	2	3						
4.	13	15	5	3						
5.	13	14	4	4					1**	
6.	1	2	6	5	4	5	2	2	5	4
7.	1	1	3	2	3	1	0	1	11	13
8.	6	3	4	5	2	3	0	1	6	6

Summary of Estimated Interrater Reliability
By Criterion Between One Pair of Raters
(Pearson Correlation Coefficients)

Criterion	Raters 1 and 2
1.	1.0000
2.	.7774
3.	.7906
4.	.7211
5.	.0311
6.	.6374
7.	.6304
8.	.7887

*An absence of data under a rating indicates a rating not described for that criterion.

**Rating given does not correspond to any rating described for criterion five. It looks like an error in marking the Answer Sheet.

Except for criterion five (Uses precise language.), these data were considered acceptable. Though comparing these coefficients directly with those resulting from the first version would be a faulty statistical procedure (after all, the criteria, raters, and directions changed), it isn't faulty at all to peek at a promising contrast, especially in the readers' agreement on ratings for the form criteria (6, 7, and 8).

Remaining to be done was an overhaul of criterion five. Why did the raters differ so markedly in applying this criterion to the writing samples? As I saw it, the only way to answer the question was to ask the raters--so I did. One SWRL rater focused on the example given in the description, i.e., "cooking room" or "room" (rather than "place"); the other rater focused on the second example given in the description, i.e., "left" or "north" (rather than "that way"). The next step was to see what would happen if we removed the examples from the description for criterion five, rewrote the description, and then had two new raters read the papers (same ones) and select a rating for this criterion only:

Criterion	Description
Uses precise language.	Most terms used are exact. Vague language is avoided. (A) Yes (B) No

The students' performance according to the two raters (SWRL staff who had not yet been involved in any phase of the writing sample project) was as follows:

Summary of Students' Performance
On Criterion Five by Two Raters (2nd Revision)

Rating			
Yes(A)		No(B)	
R1	R2	R1	R2
12	9	6	9

The interrater reliability was estimated to be .7071 (Pearson correlation coefficient), which indicated a more consistent interpretation of the criterion and its description.

Review of comments from readers involved in the revision led to a note for criterion seven that reads: "Do not count as errors the capitalization or lack of capitalization of room names (since much variation exists in the capitalization of room names on a building plan)." Three other notes were added to this criterion on the basis of recommendations for the Form B assessment. The description for criterion five was changed as noted above.

Evolution of the Writing Sample Assessment, Form B

The TEPS, English Literacy, Form B writing sample assessment began with the following note to the student, prompt, and scoring criteria*:

TRAINING AND EMPLOYMENT PREREQUISITES SURVEY

ENGLISH LITERACY 1V B

WRITING SAMPLE

NOTE TO THE STUDENT

The purpose of this writing sample is to find out what you have learned about writing. Write your name at the top of the lined paper. Then read the directions on the other side of this page. Keep the directions in front of you to remind you of the things to include when you write. After you have finished, your teacher will collect your paper and this page. You may begin.

*Criteria are numbered 111-119. These three-digit numbers correspond to the Answer Sheet, where ratings were recorded. In fact, 111 refers to criterion one, 112 to criterion two, and so forth.

DIRECTIONS: You work in the office of a large electronics company. Two weeks ago, your company ordered note pads from the Titus Print Shop. You received the shipment of note pads this morning, but they are printed with the wrong telephone number. Write a business letter to the Titus Print Shop. Tell them that you are returning the note pads and that you want them to print new ones at no additional cost to your company.

- The opening and closing parts of your letter are provided below. Write just the body of the letter.
- Tell why you are writing this letter.
- State what you want the shop to do about the problem.
- Be sure to use exact words.
- Be sure to use words that are appropriate for a business letter.

CLARKE ELECTRONICS CO., INC.

101 Industrial Way
Sunville, Arizona 85098
May 15, 1982

Titus Print Shop
19464 Main Street
McBee, Arizona 85099

Dear Sir or Madam:

Sincerely,

Andrea Lawlor

Andrea Lawlor, Manager
Supply Department

BEST COPY AVAILABLE

DIRECTIONS FOR SCORING THE WRITING SAMPLE - B

Teachers are asked to score the writing sample which students completed on the lined paper. Guidelines for scoring the writing sample are presented below. You should do your scoring directly on the back of students' answer sheets in the spaces numbered 121-129. Use a number 2 pencil to do your scoring. The scoring guide below describes three performance ratings for each criterion: good, acceptable, and unacceptable. Fill in the space marked A for good, B for acceptable, and C for unacceptable.

Most teachers will only need to read the scoring guidelines a few times before they remember the criteria for each performance rating. A simple scoring aid is provided after the guidelines.

Scoring Guide (121 through 129)

121. Submits scorable sample.

Good: (This rating is not used for this criterion.)

Acceptable: Has both of the following:

a. legible writing
b. writing relevant to directions in prompt

Unacceptable: Has any one of the following:

a. illegible writing
b. writing not relevant to directions in prompt
c. no response

CONTENT:

122. States reason for letter.

Good: Includes any two or three of the following:

a. ordering and receiving note pads
b. wrong telephone number on note pads
c. returning note pads

Acceptable: Includes any one of the following.

a. wrong telephone number on note pads
b. returning note pads

Unacceptable: Includes neither of the following:

a. wrong telephone number on note pads
b. returning note pads

183. States desired course of action.

Good: Includes both of the following:

a. print new note pads (replacements)
b. at no additional cost (free, no charge)

Acceptable: Includes: print new note pads (replacements)

Unacceptable: Does not include: print new note pads (replacements)

184. Uses precise language.

Good: Precise words are used in most or all cases to describe important features of the letter, for example:

"note pads" vs. "pads" or "things" or "stuff"
"printed with" vs. "with" or "have"
"telephone number" vs. "number" or "writing"
"additional cost" vs. "cost"

Acceptable: General words are used in most cases to describe important features of the letter. However, precise words may be used in a few cases. (See examples under Good, above.)

Unacceptable: Words used are so general that the recipient of the letter would not be able to understand the request easily, for example:

"the stuff is no good so here it is"

185. Uses language appropriate for a business letter.

Good: Has both of the following:

a. polite or neutral tone
b. impersonal tone (does not refer to person who handled order or self)

Acceptable: Has both of the following:

a. polite or neutral tone
b. personal tone (refers to person who handled order or self)

Unacceptable: Has impolite or rude tone.

1ø6. Excludes unrelated sentences.

Good: All sentences refer to situation described in prompt.

Acceptable: Most sentences refer to situation described in prompt.

Unacceptable: Most sentences do not refer to situation described in prompt.

FORM:

1ø7. Uses correct grammar and complete sentences.

Good: Has both of the following:

a. few or no grammatical errors
b. no incomplete sentences (fragments, run-ons)

Acceptable: Has any one of the following combinations:

a. few or no grammatical errors and few incomplete sentences
b. some grammatical errors and no incomplete sentences
c. some grammatical errors and few incomplete sentences

Unacceptable: Has both of the following:

a. many grammatical errors
b. many incomplete sentences

1ø8. Capitalizes and punctuates correctly.

Good: Has few or no capitalization and punctuation errors.

Acceptable: Has some capitalization and punctuation errors.

Unacceptable: Has many capitalization and punctuation errors

1ø9. Spells correctly.

Good: Has few or no misspelled words.

Acceptable: Has several different* misspelled words.

Unacceptable: Has many different* misspelled words.

*More than one instance of misspelling the same word should be evaluated as one misspelled word.

Scoring Aid (121 through 129)

Place the student's answer sheet under this page of the Manual, aligning rows 121 through 129 on the answer sheet with 121 through 129 below:

(A) Good
(B) Acceptable
(C) Unacceptable

121. Submits scorable sample.

122. States reason for letter.

123. States desired course of action.

124. Uses precise language.

125. Uses language appropriate for a business letter.

126. Excludes unrelated sentences.

127. Uses correct grammar and complete sentences.

128. Capitalizes and punctuates correctly.

129. Spells correctly.

By the time these materials were field tested for the first time, they had already undergone many revisions. Review sessions with several SWRL staff resulted in the decision to eliminate some criteria from early drafts (under CONTENT), "Identifies self and company," "Limits paragraph to one main idea," and (under FORM) "Writes legibly with appropriate margins and indentations," and to combine criteria (under CONTENT), "States purpose of letter," and "States reason for dissatisfaction" (which later became "States reason for letter"). The two content criteria were eliminated because they were considered unnecessary or inappropriate for the particular writing task. The form criterion was

withdrawn for the same reasons as it was for Form A: the first criterion already took care of legibility, margin formation could not be scored easily nor reliably, and indentation alone had limited value for assessment. The criterion formed from two was simply a move toward economy; reviewers didn't see much difference between them, given the nature of the writing task.

Participants in the first field test for the TEPS, English Literacy, Form B Writing Sample were 40 tenth-grade students (22 boys, 18 girls) who attended remedial English classes in an urban high school, their four English teachers,* and two SWRL staff. Students were given the prompt and the note shown earlier. Their responses were written in pen or pencil on lined paper. Like the students who participated in the field test for Form A, these youths all had Spanish surnames. Following are data indicating students' performance by criterion and reliability of readers' ratings (teachers and SWRL staff). In looking at the data, note that raters 1 and 2 (the teachers) assigned a rating of Good for criterion one to 19 and 11 students, respectively. The problem is that Good is not a rating option described in the scoring guide for criterion one. We can't tell whether the teachers intended to select Acceptable (instead of Good) and Unacceptable (instead of Acceptable), but followed unclear directions incorrectly, merely disregarded the scoring guidelines and fell into the customary practice of judging students' papers Good, Acceptable, or Unacceptable, or came to these ratings via some other route. If only one rater was in error, we might have simply acknowledged it as a singular occurrence. But two raters doing it warranted action.

*The four teachers divided students' papers so that each pair of teachers read half the papers. For computations, the four teachers were treated as two raters.

Summary of Students' Performance
By Criterion and Rater[a]

Criterion	Rating											
	Good(A)				Acceptable(B)				Unacceptable(C)			
	R1	R2[b]	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
1.	19	11	0	0	20	28	40	40	1	0	0	0
2.	23	26	37	34	17	13	3	6	0	0	0	0
3.	24	24	34	24	15	13	6	4	1	2	0	12
4.	24	24	16	29	11	9	24	11	5	6	0	0
5.	24	24	2	0	10	10	33	30	6	5	5	10
6.	27	29	39	37	13	9	1	3	0	1	0	0
7.	10	13	4	19	19	17	25	15	11	9	11	6
8.[c]	12	14	12	30	16	14	20	7	12	11	8	2
9.[d,e]	15	13	16	30	13	15	23	7	12	10	1	0

And the substantial differences between raters on performance judgments for many criteria, not only between the teachers and SWRL staff, but also between the two SWRL staff (the teachers tended to assign a similar number of the same ratings for individual criteria), called for a closer look at the data. What follows is a summary of estimated reliability between pairs of readers.

---

[a]Raters 1 and 2 are high school (tenth-grade) English teachers. Raters 3 and 4 are SWRL staff.

[b]Rater 2 omitted ratings for one student; hence the total number of ratings for rater 2 is 39.

[c]Rater 4 omitted a rating on criterion eight for one student; hence the total number of ratings for rater 4 for this criterion is 39.

[d]Rater 2 omitted a rating on criterion nine for two students; hence the total number of ratings for rater 2 for this criterion is 38.

[e]Rater 4 omitted a rating on criterion nine for three students; hence the total number of ratings for rater 4 for this criterion is 37.

Summary of Estimated Interrater Reliability
By Criterion Between Pairs of Four Raters
(Pearson Correlation Coefficients)

	Raters*					
Criterion	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
1.	.2001	**	**	**	**	**
2.	-.2580	.3312	.3470	.0822	.0542	.6778
3.	-.2717	-.0710	.1075	.4075	.2303	.2956
4.	-.1821	-.0433	.3347	-.0269	-.0303	.0457
5.	-.1590	.1108	.5072	.3859	.1545	.3156
6.	-.2879	-.1111	.0051	-.0908	.4114	-.0456
7.	.1546	.5192	.4467	.1578	.2209	.5483
8.	-.0410	.2766	.2249	.4509	.1347	.5014
9.	.1645	-.0072	.3248	.3116	.4878	.3605

The data above were truly discouraging; the degree of agreement between pairs of readers was so low it was difficult to believe that both readers looked at the same papers and same scoring guidelines. Since it was a fairly simple matter to recompute these figures, i.e., arrangements for field testing weren't involved, we treated the Good and Acceptable ratings as merely Acceptable, and the Unacceptable ones as they stood--we collapsed the rating options from three to two. Note the following results. (But understand that the results represent a paper manipulation and not actual ratings.) Immediately, you can see that these data were less informative in general than the data just above, i.e., with Good, Acceptable, and Unacceptable ratings; fewer coefficients could be computed for the collapsed ratings.

*Raters 1 and 2 are high school (tenth-grade) English teachers. Raters 3 and 4 are SWRL staff.

**Coefficient could not be computed because variance was zero.

Summary of Estimated Interrater Reliability
By Criterion Between Pairs of Four Raters
(Pearson Correlation Coefficients)

Collapsed Ratings

	Raters*					
Criterion	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
1.	**	**	**	**	**	**
2.	**	**	**	**	**	**
3.	-.0377	**	.2446	**	.0969	**
4.	-.1635	**	**	**	**	**
5.	-.1635	.0529	.5659	.5412	.1261	.3055
6.	**	**	**	**	**	**
7.	.0624	.3730	.2117	.3329	.2725	.3685
8.	.1994	.2182	.0969	.4492	.2751	.4971
9.	.1083	-.1048	**	-.0982	**	**

The only exit from the predicament, as we saw it, was to revise the materials and then field-test them with another group of students. For direction toward revision, readers' comments were reviewed.

Performance data, reliability data, and readers' comments led to the revised directions, writing prompt, and scoring guidelines*** below (the note to the student was retained as written):

*Raters 1 and 2 are high school (tenth-grade) English teachers. Raters 3 and 4 are SWRL staff.

**Coefficient could not be computed because variance was zero.

***Guidelines refer to criteria numbered 111-119. These three-digit numbers correspond to the Answer Sheet, where ratings were recorded. In fact, 111 refers to criterion one, 112 to criterion two, and so forth.

DIRECTIONS: You work in the office of a large electronics company. Two weeks ago, your company ordered some supplies from the Titus Office Supply Company. You received the shipment this morning, but it was not what your company had ordered. Write a business letter to the Titus Office Supply Company. Tell them what was wrong with the shipment. Tell them what you want them to do about it.

- The opening and closing parts of your letter are provided below. Write just the body of the letter.
- Tell exactly what was wrong with the shipment.
- State exactly what you want the Titus Office Supply Company to do about the shipment.
- Be sure to use
  - exact words
  - words that are appropriate for a business letter
  - good grammar and complete sentences
  - correct punctuation and spelling.

CLARKE ELECTRONICS CO., INC.

101 Industrial Way
Sunville, Arizona 85050
May 15, 1982

Titus Office Supply Co.
19464 Main Street
McBee, Arizona 85059

Dear Sir or Madam:

Sincerely,

Andrea Lawler

Andrea Lawler, Manager
Supply Department

BEST COPY AVAILABLE

Form B Guide. Spaces numbered 111 through 119 are used to describe characteristics of the writing sample for Form B. Use the descriptions below to fill in the bubbles on the student's answer sheet. Fill in only one bubble for each number. Do not fill in any bubbles if the sample is any of the following:

- illegible
- irrelevant to directions given
- not written

		(A)	(B)	(C)	(D)
111.	Estimates size of writing sample (based on 8 1/2" x 11" sheet of paper).*	Writes about a full page or more.	Writes about 3/4 page.	Writes about 1/2 page.	Writes about 1/4 page or less.

*NOTE: If handwriting is much larger or smaller than average, judge sample size according to what it would be if handwriting were of average size.

CONTENT:

112. States problem.

Mentions a specific error or problem with the shipment.

(A)	(B)
Yes	No

113. States desired course of action.

Mentions a specific action that the office supply company should take concerning the shipment.

(A)	(B)
Yes	No

114. Uses precise language.

Uses precise words to describe important points, for example:

"pens" or "desks" or "chairs" (rather than "shipment" or "things")

"wrong color" or "2 dozen" (rather than "wrong" or "not enough")

(A)	(B)
Yes	No

115. Uses language appropriate for a business letter.

Uses polite language.

(A)	(B)
Yes	No

116. Excludes unrelated sentences.

All or most sentences refer to situation described in DIRECTIONS.

(A)	(B)
Yes	No

FORM:

117. Uses correct grammar and complete sentences.

The total number of grammatical errors and incomplete sentence* errors is _____.

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: Count one error each time a segment of a run-on sentence should have been a separate sentence.

118. Capitalizes and punctuates correctly.

The total number of capitalization and punctuation errors is _____.*

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: Do not count as errors the places where a run-on sentence should have been segmented (and thus capitalized).

119. Spells correctly.

The total number of spelling errors is _____.*

(A)	(B)	(C)	(D)	(E)
none or 1	2	3	4	5 or more

*NOTE: More than one instance of misspelling the same word the same way should be evaluated as just one misspelled word.

Administration: Writing Sample

The Writing Sample is the part of the English Literacy Survey that offers an opportunity to describe characteristics of an individual's written response to a prompt. Characteristics described include content (ideas represented) and form (language usage and mechanics).

- The Writing Sample is not intended to be a timed activity. Within reasonable limits, each examinee should be given enough time (about 15-25 minutes) to complete the item.
- Answer Sheets and Survey Booklets should not be available to examinees during the administration of the Writing Sample.
- Distribute to each examinee the single sheet marked WRITING SAMPLE (Form A or B) and a sheet of lined paper.
- Have examinees read the information in the box below the words NOTE TO THE STUDENT. Examinees should be able to read independently and follow the directions for the writing sample. Be sure each examinee has written her or his name on the lined paper.
- After examinees finish writing, collect the single sheet of directions for the WRITING SAMPLE and the actual writing sample.

Directions for Describing the Writing Sample (Forms A and B)

Teachers are asked to describe the writing sample which students completed on the lined paper. Guidelines for describing the writing sample are presented below. You should use students' answer sheets and fill in the spaces enclosed in the box near the right bottom corner of SIDE ONE. This box will be marked TEACHER SCORED WRITING SAMPLE.

The guides below (Forms A and B) describe student performance for several characteristics. Refer to the guides to fill in the bubble whose description comes closest to the student's performance for each characteristic. A one-page Summary of Descriptions follows the Guide for each form. The Summary can simplify the task of describing characteristics.

Form B Summary of Descriptions. Place the student's answer sheet next to this page of the directions, to help you fill in the answer sheet quickly.

111. Sample size	(A)	(B)	(C)	(D)	
	full page +	3/4	1/2	1/4 or more/less	
112. Problem stated	(A)	(B)			
	Yes	No			
113. Action stated	(A)	(B)			
	Yes	No			
114. Precise language	(A)	(B)			
	Yes	No			
115. Polite language	(A)	(B)			
	Yes	No			
116. Related sentences	(A)	(B)			
	Yes	No			
117. Errors in grammar and incomplete sentences	(A)	(B)	(C)	(D)	(E)
	0-1	2	3	4	5+
118. Errors in capitalization and punctuation	(A)	(B)	(C)	(D)	(E)
	0-1	2	3	4	5+
119. Errors in spelling	(A)	(B)	(C)	(D)	(E)
	0-1	2	3	4	5+

Observe that the revised prompt above was written to discourage students from regurgitating the prompt itself; the prompt was designed so that responses would be more specific. Note too that students are guided to use precise language ("exact words") and business-appropriate language. The directions for administering the assessment now called for description rather than evaluation, and the structure for describing performance was built into the scoring guidelines (Yes, No for content criteria). In this

way, we hoped teachers would be freed from automatically thinking of students' writing as Good, Acceptable, and Unacceptable, and instead would attend closely to descriptions of individual criteria as they appear in the guidelines. The new first criterion (Estimates size of writing sample.) was included as a yardstick for describing elements of form (grammar and complete sentences, capitalization and punctuation, and spelling). The new form criteria, we reasoned, would give readers an easier method of describing students' performance for these elements, i.e., counting. Together, descriptions of criterion one and the form criteria would yield data that were more sensible, data that would consider number of errors per estimated unit of writing. A program for processing these data could be easily set to a standard specified by a district, school, or teacher. For example, a passing grade in spelling for a full page or more of writing could be four or fewer errors; for about 3/4 page, three or fewer; about 1/2 page, two of fewer; about 1/4 page or less, only none or one error would be passing. Further, notes at the bottom of descriptions for form criteria would, we hoped, clarify guidelines for describing troublesome situations such as run-on sentences. The Summary of Descriptions was designed to facilitate recording readers' judgments. The next step was to field-test these revised materials.

Twenty-four high school juniors and seniors who hadn't passed a suburban school district's writing sample test were administered the TEPS, English Literacy, Form B Writing Sample. All but a few students were attending a Basic Writing Skills class; a handful were in an ESL program. We have no data that further describe the population for this field test. Students' names, where given, obviously were removed from their papers; thus, we do not have information even on students' gender. The English

Department Chair and three other English teachers served as readers.* Two SWRL staff unfamiliar with both the revised and the original materials also were readers. Following the procedures of the earlier field tests, we gathered data on students' performance and on estimated interrater reliability (teachers and SWRL staff), and reviewed comments submitted by the six readers. Note first the following indication of students' performance by criterion.

Summary of Students' Performance
By Criterion and Rater[a]

	Rating[b]																			
	A				B				C				D				E			
Criterion	R1	R2	R3[c]	R4[c]	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
1.	3	4	1	2	1	1	4	2	7	7	5	6	13	12	11	11				
2.	16	11	9	8	8	13	12	13												
3.	17	15	14	17	7	9	7	4												
4.	2	6	3	4	22	18	18	17												
5.	21	21	20	19	3	3	1	2												
6.	20	19	12	18	4	4	9	3										1[d]		
7.	1	4	4	5	6	3	3	1	1	1	3	4	2	0	2	2	14	16	9	9
8.	8	6	9	10	3	6	5	4	1	4	2	3	4	0	3	1	8	8	2	3
9.	13	13	11	7	1	1	3	5	6	5	3	2	3	3	0	5	1	2	4	2

*The four teachers divided students' papers so that each pair of teachers read half the papers. For computations, the four teachers were treated as two raters.

[a]Raters 1 and 2 are high school English teachers. Raters 3 and 4 are SWRL staff.

[b]An absence of data under a rating indicates a rating not described for that criterion.

[c]Raters 3 and 4 omitted ratings for three students and commented that the three papers were unscorable; hence the total numbers of ratings for raters 3 and 4 are 21.

[d]Rating given does not correspond to any rating described for criterion six. It looks like an error in marking the Answer Sheet.

An examination of the data above shows no major rating pattern across criteria for the teachers or the SWRL staff. Agreement between ratings of the pairs of teachers and pair of SWRL staff is generally respectable except for criterion six, as shown below.

Summary of Estimated Interrater Reliability
By Criterion Between Two Pairs of Raters
(Pearson Correlation Coefficients)

	Raters*	
Criterion	1 & 2	3 & 4
1.	.9535	.9482
2.	.6504	.5095
3.	.8284	.6860
4.	.5222	.8416
5.	1.0000	.6892
6.	.3508	.1964
7.	.9667	.9632
8.	.7712	.4109
9.	.8214	.8412

The low agreement between raters on criterion six warranted close attention to readers' remarks, as did the lukewarm agreement on criterion two, four, and eight. Unfortunately, readers did not comment about criterion six. So I asked the SWRL readers to discuss their interpretation of the criterion and the way they described this characteristic of students' performance (the teachers were unavailable). The written comments received, recommendations for modifying criteria two, four, and eight, and the gist of the discussions with SWRL staff concerning criterion six and suggestions for dealing with this criterion

*Raters 1 and 2 are high school English teachers (N = 24). Raters 3 and 4 are SWRL staff (N = 21).

were used in next steps toward the final version of the TEPS writing sample assessments.

The degree of agreement between readers and their comments about criterion two recommended adjustment in the assessment materials. The question is "What gets changed?" The statements and descriptions of each criterion are only as productive as students' interpretation of the prompt that stimulates their writing and readers' interpretation of scoring materials. So the question is really twofold: Do we change the wording of the prompt?, or Do we change the wording of the statement and description of the criterion? A tentative answer lies in readers' comments--in those remarks that address the prompt in general, as well as in those that relate to criterion two specifically. A reasonable modification would be in the wording of the prompt, something along the lines of a comment made by one of the English teachers, i.e.,

> Choose a specific item(s) that had been ordered.
>
> Tell specifically what was wrong with the item(s) sent and why it should be changed: name problem and change.

The expectation is that readers will be able to describe students' statement of the problem more easily if greater specificity is induced in the writing sample.

Criterion four might also be improved with a reworded prompt. Even though SWRL staff described the papers similarly for this criterion (the coefficient was .8416), the teachers were not in such close agreement (.5222). And comments from both sets of readers indicated that this criterion was troublesome. A tentative recommendation is to observe the results after the prompt is reworded to inform the student clearly that specific details should be improvised.

Based on uneven and somewhat low coefficients of estimated interrater reliability, the third criterion that seemed to warrant special attention was eight (agreement between teachers, .7712; but between SWRL staff, only .4109). SWRL staff did not comment on this criterion, and the only remark made by teachers was brief: "Run-ons fit here." An examination of the ratings student by student revealed a curious phenomenon: when the pair of raters agreed on criterion eight (10 out of 21 scorable papers), the most frequent rating given (7 papers) was (A) none or 1 [capitalization and punctuation error]; of the other three cases of agreement, two were (B) 2 [capitalization and punctuation errors] and only one was (E) 5 or more. In other words, raters were more likely to agree when they described papers as having few capitalization and punctuation errors. Failing to locate sources of disagreement by examining the papers themselves, I asked the two SWRL raters to discuss their ratings. The meeting between the three of us was informative. It disclosed an occasional oversight of the note that told raters not to count as errors the points where a run-on sentence should have been segmented (and thus capitalized). The meeting also enabled us to see the need for informing raters whether the letter's address, greeting, and closing were to be included in the assessable writing (they weren't, since these were provided in the prompt). Three other outcomes of the meeting were recommendations to advise raters that the same capitalization or punctuation error should be counted only once, that sentence fragments introduced by a capital should not be counted as errors (they'd already be counted under criterion seven), and to provide examples of capitalization and punctuation errors that could be overlooked easily (and were

by the SWRL raters), for example, capitals within words, hyphens, apostrophes, and terminal periods.

Criterion six (Excludes unrelated sentences) yielded the lowest rating correlations among all criteria, both for teachers and for SWRL staff (.3508 and .1964, respectively). Yet no reader commented about it. Examining students' performances and actual papers and talking individually with the SWRL readers (the teachers were unavailable), I learned that two widely different interpretations of the criterion probably led to rating discrepancies. One reader held a literal interpretation of the word refer in the description of the criterion, All of most sentences refer to situation described in DIRECTIONS. That is, if a sentence did not include words that were specific to the prompt, that sentence was counted among the ones that didn't meet the description--they didn't refer to the situation. The other reader interpreted refer more loosely, essentially as "having to do with." For example, a paper included these sentences: "The shipment we recieved [sic] was of no use to us. it [sic] was for the industrial building next door from us." Another paper had this sentence: "This may have come from one of your employees wrong." One reader judged that these sentences did not refer to the situation described in the prompt. The other reader, however, thought the sentences did refer to the general situation of a letter of request for an order adjustment. Similarly, one reader viewed as unrelated to the prompt those sentences that expressed amenities of business correspondence; the other reader thought such sentences were related, for instance, "I am very sorry to write this kind of letter" and "Thank you for helping [sic]."

To eliminate the confusion that might arise from the negative criterion statement Excludes unrelated sentences, it was changed to Stays on topic. Also to reflect communication that crosses sentence boundaries

and to avoid the wide interpretation typically given to the word most (some people think of most as 51 percent or more, others think of it as closer to 80 percent or so), the criterion description was reworded as Generally relates to situation described in DIRECTIONS (this revision would also be expected to eliminate the problems of interpretation associated with the word refer). Another recommendation was to include a note in the scoring guidelines that amenities of business correspondence do not constitute abandonment of the topic. I had some concern that if the suggestions above were incorporated in a revision, the resulting rating would be Yes for all or nearly all papers, since papers that were truly irrelevant to the topic were to be left unscored. A criterion that provides consistently uniform ratings would be wasteful. Why not just describe the performance without rating papers for it? But this field test demonstrated that teachers may not judge a paper unscorable as readily as other raters (i.e., SWRL staff) would. In fact, three papers judged unscorable by SWRL staff, for content irrelevant to the prompt, were indeed scored by teachers. Could the reason be that teachers are in need of information concerning their own students' writing performance and are thus more inclined to retain a paper for further scoring? In administrations of the assessment in its final form, teachers will be the readers and raters. Hence, this line of reasoning advised suspending, at least for now, the concern regarding uniform rating on criterion six.

Although we would like to think that the assessments are "final," that would be premature. Both writing sample assessments will be subject to further revision based on additional experience with them.

As with any written communication, a writing assessment can be dashed off in one draft and it can also be revised to death. Reasonable effort is between those extremes. Reader reaction is the best gauge of what to do after the first draft. Because the "readers" of a writing assessment react in writing, the reaction is unambiguous. All the "author" has to do is to respond to the reactions that are registered. Doing that is straightforward, but it's not as simple as it seems.

References

Escoe, A. Remedial reading and English instruction for junior high and and high school students: A description of common practices in teaching and assessment, (TR 75). Los Alamitos, CA: SWRL Educational Research and Development, 1982(a)

Escoe, A. Skills index and technical characteristics of instrumentation for describing youth training and employment prerequisites in English literacy, (Technical Note No. 2-82/10). Los Alamitos, CA: SWRL Educational Research and Development, 1982(b)

Gentry, L. Entry-level occupations and major employers in Southern California, (Technical Note No. 2-82/11). Los Alamitos, CA: SWRL Educational Research and Development, 1982.

Humes, A. A method for evaluating writing samples, (Technical Note No. 2-80/02). Los Alamitos, CA: SWRL Educational Research and Development, April 1980.

Humes, A., Cronnell, B., Lawlor, J., Gentry, L., & Fieker, C. Specifications for 1-6 writing skills assessment, (Technical Note No. 2-80/18). Los Alamitos, CA: SWRL Educational Research and Development, August 1980.

Nack, S. Basic skills viewed by employers as necessary for entry level jobs in service related industries, (Technical Note No. 2-82/09). Los Alamitos, CA: SWRL Educational Research and Development, 1982.